THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 7, September 1981
Printed in U.S.A.

A Comparison of Three Speech Coders to be Implemented on the Digital Signal Processor

By R. V. COX

(Manuscript received June 24, 1980)

The recently developed digital signal processor is a device used for implementing low- to medium-complexity speech coders. It is currently being used in implementing adaptive differential pulse-code modulation (ADPCM) coding, two-band sub-band coding, and four-band sub-band coding. This study was performed to determine optimal parameter values for the two sub-band coders in preparation for their implementation on the digital signal processor and to determine their performance relative to ADPCM. (The actual implementation of the ADPCM and two-band sub-band algorithms are discussed in other papers in Part 2 of this issue of the Bell System Technical Journal.*) Performance was judged on the basis of segmental signal-to-noise ratio and a forced-choice, subjective comparison test of the coders. All three coders were simulated at bit rates of 16, 20, 24, 28, and 32 kb/s. The simulations were performed on a laboratory computer.*

I. INTRODUCTION

The recently developed DSP is a device for implementing low-to medium-complexity speech coders. Three coders are currently being implemented. The simplest coder is adaptive differential pulse-code modulation (ADPCM) and is discussed in Ref. 1. The other two are in the sub-band coder (SBC) family. Of these, the simpler one is two-band sub-band coding (2B-SBC), featuring quadrature mirror filtering and two equal bands. It is discussed in Ref. 2. The other coder—the most complicated—is four-band sub-band coding (4B-SBC), featuring four equal bands. Its implementation is still in progress.

This report discusses the initial design parameters for the latter two coders and the relative performance of all three. Segmental signal-to-noise ratio (SNR) measurements were made on all three coders via computer simulation at five different bit rates, 16, 20, 24, 28, and 32

kb/s. In addition, 12 subjects ranked the coders in a comparison test. The simulations reported here were carried out on a laboratory computer as preparation for the implementation of the SBC coders on the DSP.

Section II reviews the design of ADPCM and discusses the design of the two sub-band coders. Section III discusses the results of the subjective testing experiment, and Section IV gives the conclusions of this study.

II. DESIGN OF THE CODERS

2.1 Design of ADPCM

The ADPCM design simulated here is based on the design of Cummiskey et al.[3] A block diagram of the ADPCM coder described below is shown in Fig. 1. The most significant change from the design in Ref. 3, is that only two multiplier values are used in changing the step-size, regardless of the bit rate. This is based on the ADPCM implemented by Johnston and Goodman.[4] This version of ADPCM has already been

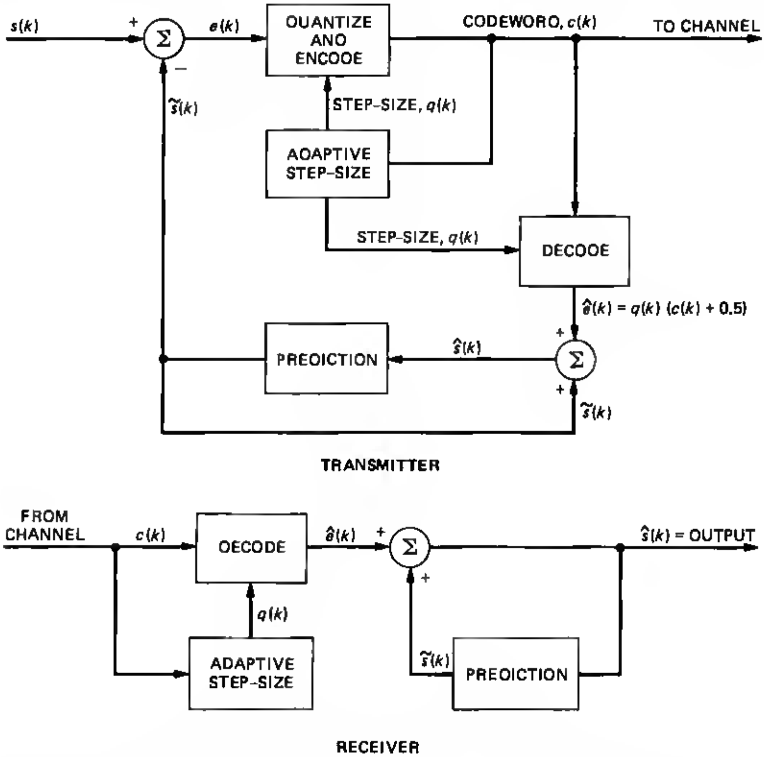

Fig. 1—Adaptive differential pulse-code-modulation coder used in simulation.

implemented on the DSP and is described in Ref. 1. The ADPCM design was also used for quantizing the sub-band signals in the other coders.

To simulate 20 kb/s and 28 kb/s ADPCM, alternating quantizers were used. For 20 kb/s, the two quantizers used are for 2 and 3 bits. Since the step-size is adapted based only on the most significant magnitude bit, the same step-size adaptation algorithm is used for all samples. The ratio of 2- to 3-bit quantizer step-size is held constant. This requires one additional multiplication to convert from the 2- to 3-bit step-size.

2.2 Two-band SBC

All sub-band coders are made from a few fundamental building blocks. The first is linear filtering to divide the signal into two or more sub-bands. These sub-bands can then be decimated to a lower sampling rate than the original signal. Some form of quantization must be used to encode and quantize each band. Interpolation and additional linear filtering is used to bring each band back to the original sampling rate and to its original space in the frequency spectrum. At this point, they can be added together to produce an output signal.

The quadrature mirror filtering technique is fairly well known for its use with sub-band coders. Each pair of quadrature mirror filters (QMFs) produces two sub-bands of equal width in frequency. Johnson has compiled a collection of different length QMFs.[5] The possible quantizers which can be used are adaptive delta modulation (ADM), ADPCM, and adaptive pulse-code modulation (APCM). Each of these techniques is fairly well known and understood. Likewise interpolation and decimation are also well understood. So what remains is the task of combining these building blocks in such a way as to fit on the DSP and, also, give the best possible performance. One of the tasks of this study was to choose good candidates for implementation.

The 2B-SBC design is based on the 2B-SBC commentary grade coder of Johnston and Crochiere.[6] That coder was developed with the object of maintaining a high-quality AM radio signal. Its parameters were tuned to music rather than to speech. This section describes parameters for a speech bandwidth version. There are five possible bit rates envisioned. For a more detailed discussion of sub-band coding in general and the exact implementation of this coder refer to Ref. 2.

Figure 2 is a block diagram of the 2B-SBC. The input speech has been band-limited from 200 to 3200 Hz by a sharp bandpass filter and sampled at 8000 Hz. A 32-tap QMF designed by Johnston is used for separating the digitized speech into the two sub-bands.[5] After 2-to-1 decimation on both bands, we found average correlations for speech of 0.7 and −0.45 for the low and high bands, respectively.

The two bands are then coded using either ADPCM or ADM. The ADM

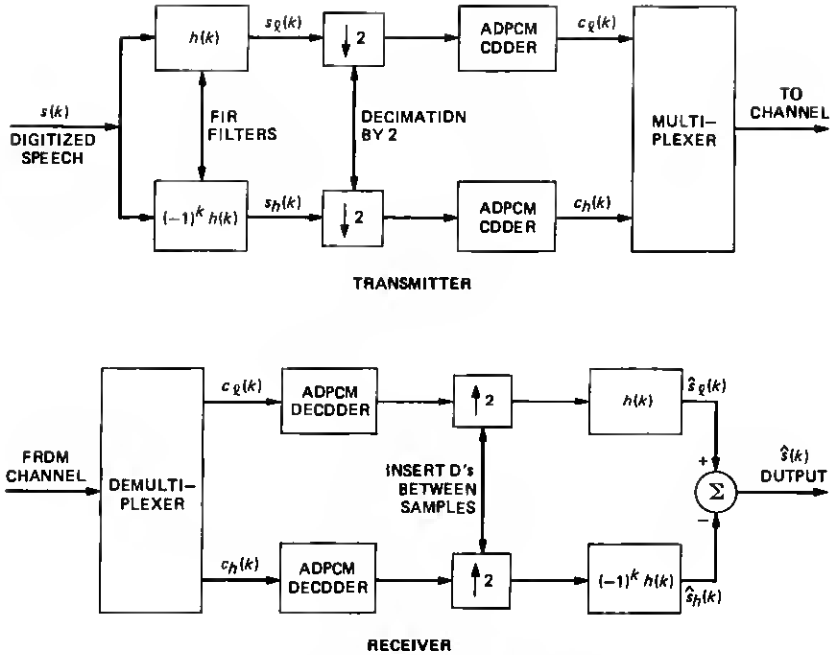

Fig. 2—Two-band sub-band coder used in simulation.

is used only for the higher band at low bit rates. It is based on the ADM of Jayant.[7] The prediction coefficients used for these coders were the correlation values mentioned above. Since the high band has a negative correlation, it was frequency inverted before quantization by the ADM, because this ADM requires a positive correlation for its adaptation mechanism to work properly. Since frequency inversion just means changing the sign of every other sample, this is a very minor operation.

The next step was to determine optimal bit allocations for the low and high bands. After experimenting with different bit allocations and evaluating them on the basis of segmental SNR measurements and informal listening, the following bit allocations were adopted for the five bit rates:

16 kb/s low: 3 bits high: 1 bit (ADM)
20 kb/s low: 4 bits high: 1 bit (ADM)
24 kb/s low: 5 bits high: 1 bit (ADM)
28 kb/s low: 5 bits high: 2 bits
32 kb/s low: 5 bits high: 3 bits.

Some alternative designs were very close. For instance, a (4, 2) allocation for 24 kb/s is almost as good as (5, 1) for the speech it was tested on. Perhaps if the speech were less sharply bandpass-filtered

and if there were more high-frequency content (such as in telephone speech) the better allocation would be (4, 2) for 24 kb/s.

2.3 Four-bend SBC design

The 4B-SBC design described here is new, although it is a logical extension of the 2B-SBC mentioned above. It starts with the same two sub-bands as the two-band design. Both of these bands are then divided into two new bands, yielding a total of four equally spaced bands. The filter used for the additional division in each band is the 16-tap QMF of Johnston designated C in Ref. 5. Figure 3 shows a block diagram for this coder.

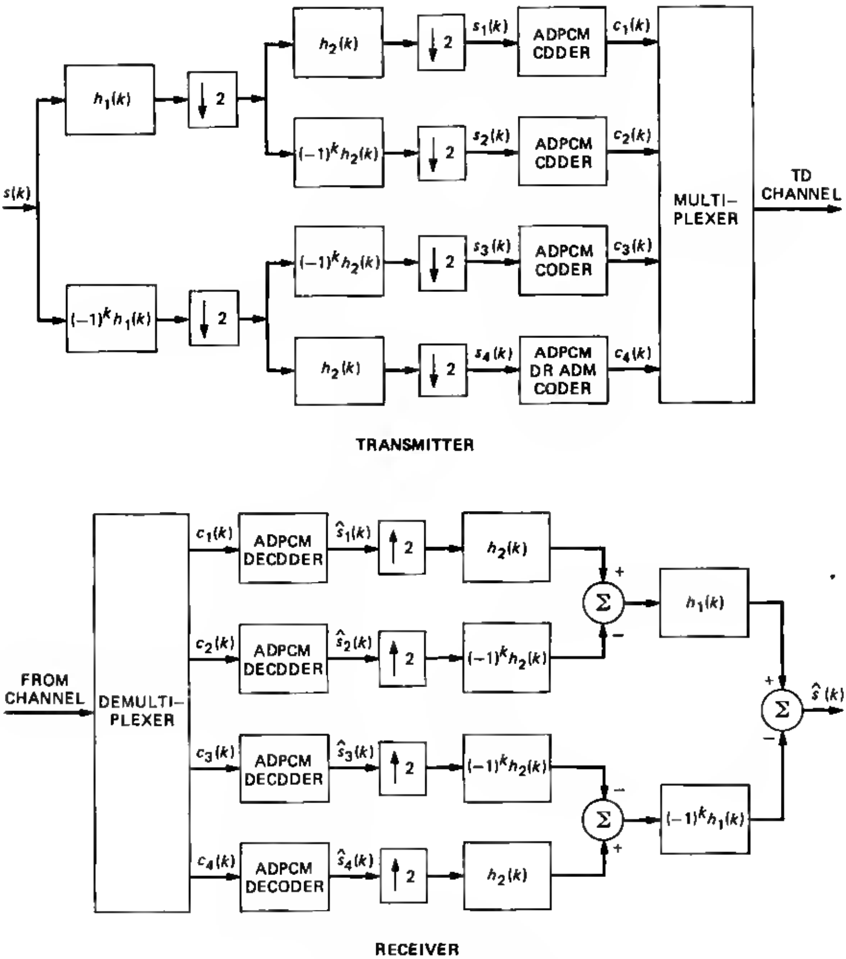

Fig. 3—Four-band sub-band coder used in simulation.

Once more the bands are quantized using ADPCM or ADM. Our measurements of average correlation for speech data showed correlations of 0.4, 0, 0, and 0.8 for the four bands going from lowest to highest in frequency. The fourth band (3000 to 4000 Hz) has actually been bandpass-filtered to cut off at 3200 Hz. As a result, it contains little power and can be ignored for low-bit rate coders. The correlations of the two middle bands are zero, reflecting that the long-term average of the speech spectrum from 1000 to 3000 Hz is flat. If a prediction coefficient of zero is used with ADPCM, the result is APCM. Thus, the two middle bands are APCM-encoded. The largest amount of power is in the first band; therefore, it receives the most bits.

The bit allocations found to be the best by the same segmental SNR measurements and casual listening process were as follows:

16 kb/s 4,2,2,0 (bands 1 to 4)
20 kb/s 5,2,2,1 (ADM on band 4)
24 kb/s 5,4,2,1
28 kb/s 7,4,2,1
32 kb/s 7,4,3,2.

The greatest amount of error occurs in the lowest band. Even at the high rates (28 and 32 kb/s) this error is still perceptible as a low rumbling noise. However, it was found that a high-pass filter with a cutoff of 200 Hz eliminated this problem. The filter used was a 121-tap FIR filter. Table I gives the coefficients, and Fig. 4 shows the frequency response. A much smaller IIR filter could also be used to do the same job.[2] Note that the above bit assignments were made without using the FIR filter. With a high-pass filter, fewer bits could be allocated to the lowest band and more to bands two and three.

2.4 Relative complexity of designs

The ADPCM designs for 16, 24, and 32 kb/s have already been implemented on the DSP.[1] The combined encoder and decoder algorithms use 48 percent of the DSP real-time capability for a sampling

Table I—Coefficients for symmetric FIR high-pass filter

−.153203E-01	−.488296E-03	−.427259E-03	−.305185E-03	−.152593E-03
.610370E-04	.305185E-03	.610370E-03	.946074E-03	.131230E-02
.170904E-02	.213630E-02	.259407E-02	.305185E-02	.350963E-02
.396740E-02	.442518E-02	.485244E-02	.524918E-02	.558488E-02
.589007E-02	.610370E-02	.625629E-02	.631733E-02	.628681E-02
.616474E-02	.592059E-02	.555437E-02	.509659E-02	.448622E-02
.37537BE-02	.283822E-02	.183111E-02	.732444E-03	−.549333E-03
−.192267E-02	−.344859E-02	−.506607E-02	−.677511E-02	−.857570E-02
−.104679E-01	−.123905E-01	−.144047E-01	−.164190E-01	−.184332E-01
−.204779E-01	−.224921E-01	−.244453E-01	−.263680E-01	−.281991E-01
−.299387E-01	−.315867E-01	−.331126E-01	−.344554E-01	−.356761E-01
−.367443E-01	−.375988E-01	−.383007E-01	−.387890E-01	−.390942E-01
.100000E-01				

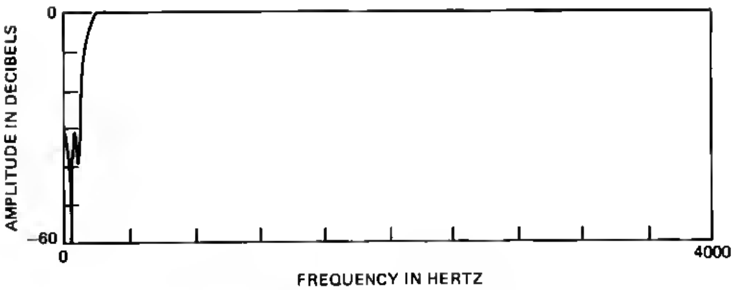

Fig. 4—Frequency response of high-pass filter for 4B-SBC.

rate of 8 kHz. An even lesser percentage of RAM and ROM memory is used. The 2B-SBC based on the design parameters reported here has also been implemented on the DSP chip.[2] It uses 98 percent of the real-time capability and 78 percent of the RAM memory. It includes an IIR bandpass filter for the input. The 4B-SBC algorithm is planned for implementation in the near future. Since all of the major portions of the 4B-SBC have been programmed already for the 2B-SBC implementation, it is possible to project how much of the DSP will be used. Both the transmitter and the receiver will require a DSP and each will use about the same fractions of real-time capability and RAM as the complete 2B-SBC algorithm. Therefore, so we might classify the three coders as having complexities of 0.5, 1, and 2, respectively.

III. RELATIVE PERFORMANCE OF THE CODERS

Since the SBC designs are more complex, a demonstration of their improved performance over ADPCM was needed to justify their implementation on the DSP. To demonstrate their relative performance all three coders were simulated on a laboratory computer. Each processed speech from a stored file. The results were evaluated by both an objective and subjective measure. The objective measure was segmental SNR, while the subjective measure was a forced-choice, subjective (A–B) comparison test in which all possible coders were compared.

Six phonetically balanced sentences were used for evaluating the coders. Three were spoken by male speakers and three by females. They were recorded using a linear microphone. They were band-limited from 200 to 3200 Hz and sampled at 8000 Hz using a 15-bit linear quantizer.

3.1 Segmental signal-to-noise ratio results

In computing segmental SNR measurements, blocks of speech of 32 ms were used. The ADPCM coder was compared with the original input speech. The SBC coders were compared with reassembled speech which had been processed by the appropriate QMF filtering, but with no

quantization. These slightly modified speech signals cannot be distinguished from the original in casual listening. Without them it would be difficult to make a fair comparison of the three coders on the basis of SNR. The measurements on 4B-SBC were made before the 121-tap FIR high-pass filtering.

The results of these measurements are summarized in Fig. 5. They show that the more complex SBC coders have a definite advantage over ADPCM at the lower-bit rates. Interestingly at 32 kb/s, ADPCM beats both of the more complex coders. The 4B-SBC maintains a fairly constant 2-dB advantage over 2B-SBC. In terms of bit rate this translates to 4 kb/s. At the low rates, the 4B-SBC has about a 6-kb/s advantage over ADPCM.

3.2 Subjective testing of the three coders

An A–B comparison test was performed to rank the three coders. Each coder at each rate was compared twice against every other coder at every rate, as well as against the original. In the two comparisons of the two coders, each one was played in first position once. There were 12 participants in the test and altogether tbere were 240 comparisons. The test was broken down into two parts, one with 110 comparisons, the other with 130. The participants listened over headphones in a soundproof booth. The participants were also broken down into two groups of six. If one group listened to a particular A–B comparison with a female speaker the other group heard a sentence with a male speaker and vice versa. Thus, we attempted to make a totally balanced and unbiased test.

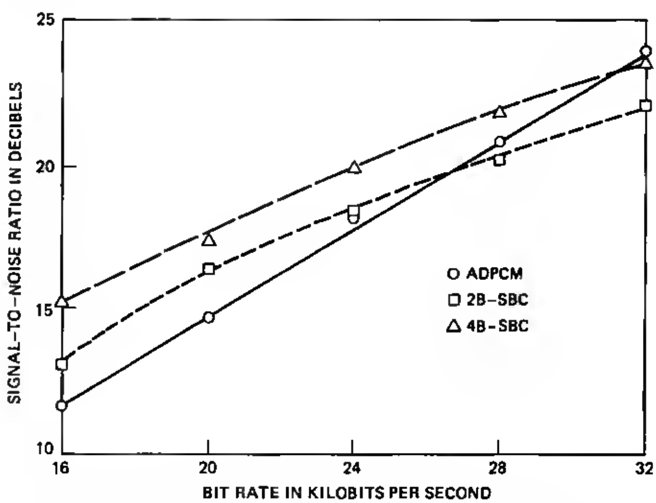

Fig. 5—Segmental SNR measurements for three coders.

Table II—Coder versus coder ratings
(Number represent percent obtained by coder listed at left of line against other coders)

		ADPCM					2B-SBC					4B-SBC				
	Orig	16	20	24	28	32	16	20	24	28	32	16	20	24	28	32
ADPCM																
16	0	—	17	0	0	0	33	8	0	0	0	0	0	0	0	0
20	0	83	—	4	0	0	79	8	4	0	0	8	0	4	0	0
24	0	100	96	—	4	21	92	63	17	13	17	75	21	21	8	13
28	8	100	100	96	—	21	96	63	67	54	33	58	63	21	25	13
32	25	100	100	79	79	—	100	87	100	63	50	100	83	58	50	54
2B-SBC																
16	0	67	21	8	4	0	—	4	0	0	0	8	0	0	0	0
20	0	92	92	37	37	13	96	—	33	8	13	37	29	13	8	13
24	4	100	96	83	33	0	100	67	—	25	37	58	50	37	29	13
28	8	100	100	87	46	37	100	92	75	—	50	71	63	33	42	29
32	17	100	100	83	67	50	100	87	63	50	—	100	96	58	46	29
4B-SBC																
16	8	100	92	25	42	0	92	63	42	29	0	—	42	13	4	4
20	4	100	100	79	37	17	100	71	50	37	4	58	—	29	13	0
24	4	100	96	79	79	42	100	87	63	67	42	87	71	—	46	25
28	21	100	100	92	75	50	100	92	71	58	54	96	87	54	—	25
32	37	100	100	87	87	46	100	87	87	71	71	96	100	75	75	—

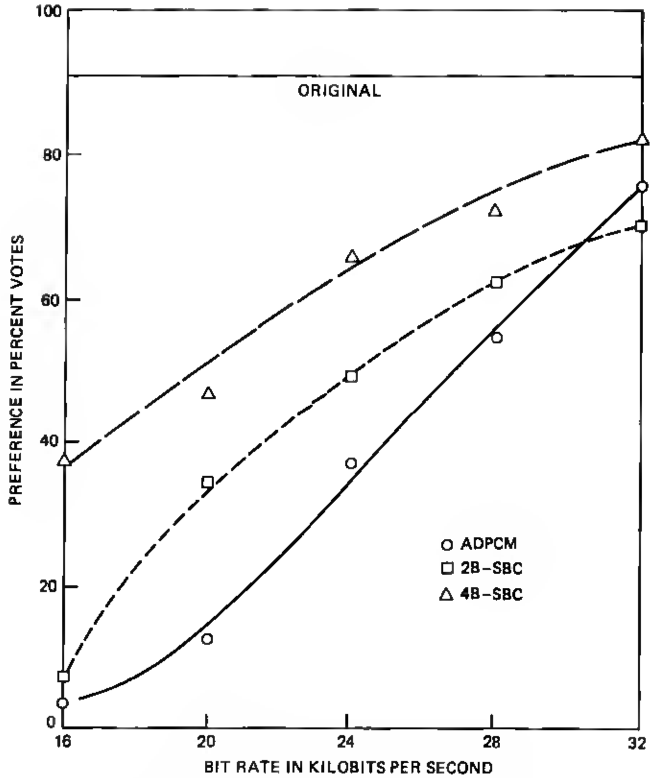

Fig. 6—Overall preference rankings for three coders.

Table II gives the individual coder versus coder comparisons. In addition, an overall preference ranking was computed based on the total number of votes received by each coder. In all, a total of 360 votes could be received by any coder. Figure 6 shows the percentage of the 360 possible votes received for each coder. This result is in good agreement with the results of Fig. 4. For example, both sub-band coders show adantage over ADPCM at the low rates and ADPCM catches up or passes them at the high rates.

Some of the more significant results are the following:

(*i*) The 4B-SBC has an 8-kb/s perceptual advantage over ADPCM at the low rates. The 24-kb/s ADPCM has been used for voice storage and playback systems.[8] This result indicates that 16-kb/s 4B-SBC could be substituted at a 33 percent savings in storage or, equivalently, a 50 percent increase in message storage capability. Moreover, at 20-kb/s, 2B-SBC has a 4-kb/s perceptual advantage over ADPCM.

(*ii*) Although 4B-SBC lost to ADPCM at 32 kb/s in SNR measurements, it beat ADPCM in the subjective tests. In addition, in direct

comparisons with the original, 32-kb/s 4B-SBC received 37.5 percent of the votes, an almost equipreferential rating. This indicates it is high quality. Since 32-kb/s ADPCM is often described as toll quality, then 32-kb/s 4B-SBC also deserves this ranking.

(*iii*) The 2B-SBC seems to provide a good alternative to ADPCM at the low bit rates for a modest increase in complexity.

IV. CONCLUSIONS

We have presented measurement data and simulation results for use in implementing two sub-band coders on the DSP. This data has already been used to design and implement the 2B-SBC and is being used for a planned 4B-SBC implementation. Simulations of these candidate coders were made on a laboratory computer. The results of these simulations indicate 2B-SBC and 4B-SBC have important advantages over ADPCM at low bit rates. This advantage is as much as 8 kb/s for 4B-SBC and 4 kb/s for 2B-SBC. The 16-kb/s 4B-SBC could be substituted for 24-kb/s ADPCM in a voice storage and playback system. In addition, 4B-SBC is rated as better quality at 32 kb/s than at 32-kb/s ADPCM.

Since the complexity of these coders is within an order of magnitude of ADPCM they should be considered as viable alternatives.

REFERENCES

1. J. R. Boddie et al., "Digital Signal Processor: Adaptive Differential Pulse-Code-Modulation Coding," B.S.T.J., this issue, Part 2.
2. R. E. Crochiere, "Digital Signal Processor: Sub-band Coding," B.S.T.J., this issue, Part 2.
3. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive quantization in differential PCM coding of speech," B.S.T.J., *52* (September 1973), pp. 1105–18.
4. J. D. Johnston and D. J. Goodman, "Multipurpose hardware for digital coding of audio signals," IEEE Trans. on Comm., *COM-26* (November 1978), pp. 1785–8.
5. J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," Proc. Int. Conf. Acoustics, Speech and Signal Processing, 1980 (April 1980), pp. 291–4.
6. J. D. Johnston and R. E. Crochiere, "An all digital 'commentary grade' sub-band coder," J. Audio Engineering Society, *27*, No. 11 (November 1979), pp. 855–65.
7. N. S. Jayant, "Digital coding of speech waveforms: PCM, DPCM and DM quantizers," Proc. IEEE, *62* (May 1974), pp. 611–32.
8. L. H. Rosenthal et al., "A multiline computer voice response system using ADPCM coded speech," IEEE Trans. on Acoustics, Speech and Signal Processing, *ASSP-22* (October 1974), pp. 339–52.